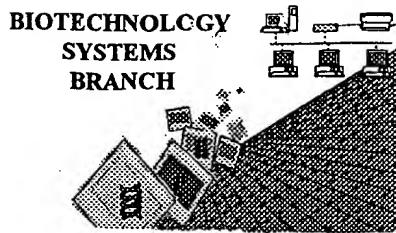1636

RAW SEQUENCE LISTING ERROR REPORT

H
42

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: _08/487,032 β_

Source: _/600_

Date Processed by STIC: _11/27/2002_

RECEIVED

DEC 1 0 2002

TECH CENTER 1600/2900

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.
PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:
1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY
FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.
PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)
PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE **CHECKER VERSION 3.1 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:
        http://www.uspto.gov/web/offices/pac/checker

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.
Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.
Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<http://www.uspto.gov/ebc/efs/downloads/documents.htm> , EFS Submission User Manual - ePAVE)
2. U.S. Postal Service: U.S. Patent and Trademark Office, Box Sequence, P.O. Box 2327, Arlington, VA 22202
3. Hand Carry directly to:
   U.S. Patent and Trademark Office, Technology Center 1600, Reception Area, 7th Floor, Examiner Name, Sequence Information, Crystal Mall One, 1911 South Clark Street, Arlington, VA 22202
        Or
   U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202
4. Federal Express, United Parcel Service, or other delivery service to: U.S. Patent and Trademark Office, Box Sequence, Room 1B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

Revised 01/29/2002

1600

RAW SEQUENCE LISTING   DATE: 11/27/2002
PATENT APPLICATION: **US/08/487,032B** TIME: 09:12:44

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

RECEIVED
DEC 10 2002
~H CENTER 1600/290

SEQUENCE LISTING
3 (1) GENERAL INFORMATION:
5 (i) APPLICANT: DOUGLAS SMITH
7 (ii) TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES
8   RELATING TO HELICOBACTER PYLORI FOR
9   DIAGNOSTICS AND THERAPEUTICS
11 (iii) NUMBER OF SEQUENCES: 941
13 (iv) CORRESPONDENCE ADDRESS:
14   (A) ADDRESSEE: LAHIVE & COCKFIELD, LLP
15   (B) STREET: 28 State Street
16   (C) CITY: Boston
17   (D) STATE: Massachusetts
18   (E) COUNTRY: USA
19   (F) ZIP: 02109
21 (v) COMPUTER READABLE FORM:
22   (A) MEDIUM TYPE: Floppy disk
23   (B) COMPUTER: IBM PC compatible
24   (C) OPERATING SYSTEM: PC-DOS/MS-DOS
25   (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
27 (vi) CURRENT APPLICATION DATA:
28   **(A) APPLICATION NUMBER: US/08/487,032B**
29   **(B) FILING DATE: 07-Jun-1995**
31 (viii) ATTORNEY/AGENT INFORMATION:
32   (A) NAME: Mandragouras, Amy E.
33   (B) REGISTRATION NUMBER: 36,207
34   (C) REFERENCE/DOCKET NUMBER: GTN-001
36 (ix) TELECOMMUNICATION INFORMATION:
37   (A) TELEPHONE: (617)227-7400
38   (B) TELEFAX: (617)227-5941

Does Not Comply
Corrected Diskette Needed

ERRORED SEQUENCES

18782 (2) INFORMATION FOR SEQ ID NO: 458:
18784 (i) SEQUENCE CHARACTERISTICS:
18785   (A) LENGTH: 122 amino acids
18786   (B) TYPE: amino acid
18787   (D) TOPOLOGY: linear
18789 (ii) MOLECULE TYPE: protein
18791 (iii) HYPOTHETICAL: YES
18793 (vi) ORIGINAL SOURCE:
18794   (A) ORGANISM: Helicobacter pylori
18796 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 458:

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

```
18798 Met Gln Ala Leu Lys Ser Leu Leu Glu Val Ile Thr Lys Leu Gln Asn
18799 1               5                   10              15
18801 Leu Gly Gly Tyr Leu Met His Ile Ala Ile Phe Ile Ile Phe Ile Trp
18802             20              25              30
18804 Ile Gly Gly Leu Lys Phe Val Pro Tyr Glu Ala Glu Gly Ile Ala Pro
18805             35              40              45
18807 Phe Val Xaa Asn Ser Pro Phe Phe Ser Phe Met Tyr Lys Phe Glu Lys
18808         50              55              60
18810 Pro Ala Tyr Lys Gln His Lys Met Ser Glu Ser Gln Ser Met Gln Glu
18811 65              70              75              80
18813 Glu Met Gln Asp Asn Pro Lys Ile Val Glu Asn Lys Asn Gly Ile Lys
18814             85              90              95
18816 Lys Thr Ala Leu His Leu Val Ala Glu Gly Leu Gly Ile Thr Ile Met
18817             100             105             110
18819 Ile Leu Gly Ile Leu Val Leu Leu Gly Leu
18820         115             120
19501 (2) INFORMATION FOR SEQ ID NO: 473:
19503     (i) SEQUENCE CHARACTERISTICS:
19504         (A) LENGTH: 203 amino acids
19505         (B) TYPE: amino acid
19506         (D) TOPOLOGY: linear
19508 *  (ii) MOLECULE TYPE: protein
19510    (iii) HYPOTHETICAL: YES
19512     (vi) ORIGINAL SOURCE:
19513         (A) ORGANISM: Helicobacter pylori
19515     (ix) FEATURE:
19516         (A) NAME/KEY: misc_feature
19517         (B) LOCATION: 1...203
19518         (D) OTHER INFORMATION: /note= "flagellar biosynthesis protein flha"
19520     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 473:
19522 Val Lys Lys Tyr Ala Glu Asp Phe Ile Thr Lys Asp Glu Val Lys Ser
19523 1               5                   10              15
19525 Leu Leu Glu Arg Leu Ala Lys Asp Tyr Pro Thr Ile Val Glu Glu Ser
19526             20              25              30
19528 Lys Lys Ile Pro Thr Gly Ala Ile Arg Ser Val Leu Gln Ala Leu Leu
19529         35              40              45
19531 His Glu Lys Ile Pro Ile Lys Asp Met Leu Thr Ile Leu Glu Thr Ile
19532     50              55              60
19534 Thr Asp Ile Ala Pro Leu Val Gln Asn Asp Val Asn Ile Leu Thr Glu
19535 65              70              75              80
19537 Gln Val Arg Ala Arg Leu Ser Arg Val Ile Thr Asn Ala Phe Lys Ser
19538             85              90              95
19540 Glu Asp Gly Arg Leu Lys Phe Leu Thr Phe Ser Thr Asp Ser Glu Gln
19541             100             105             110
19543 Phe Leu Leu Asn Lys Leu Arg Glu Asn Gly Thr Ser Lys Ser Leu Leu
19544         115             120             125
19546 Leu Asn Val Gly Glu Leu Gln Lys Leu Ile Glu Gly Val Ser Glu Gly
19547         130             135             140
19549 Ala Met Lys Val Leu Gln Lys Gly Ile Ala Pro Val Ile Leu Ile Val
```

145 ← insert

110 ← insert these numbers (do not use TAB codes between amino acid numbers. Use space characters.)

E↑ 19550 145   150   155   160
19552 Glu Pro Asn Leu Arg Lys Ala Leu Ser Lys Gln Met Glu Gln Ala Arg
19553   165   170   175
19555 Asn Asp Gly Leu Val Leu Ser His Ala Glu Leu Asp Pro Asn Ser Asn
19556   180   185   190
19558 Phe Glu Ala Leu Gly Thr Ile His Ile Asn Phe
19559   195   200
20574 (2) INFORMATION FOR SEQ ID NO: 490:
20576   (i) SEQUENCE CHARACTERISTICS: →) 211 shown
20577   (A) LENGTH: (166) amino acids
20578   (B) TYPE: amino acid
20579   (D) TOPOLOGY: linear
20581   (ii) MOLECULE TYPE: protein
20583   (iii) HYPOTHETICAL: YES
20585   (vi) ORIGINAL SOURCE:
20586   (A) ORGANISM: Helicobacter pylori
20588   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 490:
20590 Val Gly Val Ile Tyr Ile Val Thr Thr Asn Thr Leu Asn Ile Leu Ser
20591 1   5   10   15
20593 Cys Glu Ser Phe Glu Ile Leu Glu Lys Arg Glu Leu Asp Thr Ser Gly
20594   20   25   30
20596 Val Thr Lys Thr Ser Thr Pro Phe Phe Ser Arg Val Glu Gly Ile Asp
20597   35   40   45
20599 Ala Gly Thr Leu Gly Lys Leu Phe Ser Gly Ser Gln Ser Lys Asn Tyr
20600   50   55   60
20602 Phe Ala Tyr Tyr Asp Ala Leu Val Lys Lys Glu Lys Arg Lys Glu Val
20603 65   70   75   80
20605 Arg Ile Glu Lys Lys Glu Glu Arg Ile Asp Ala Arg Glu Asn Lys Arg
20606   85   90   95
20608 Glu Ile Lys Gln Glu Ala Ile Lys Glu Pro Lys Lys Ala Asn Gln Gly
20609   100   105   110
W↑→ 20611 Thr Glu Asn Ala Pro Thr Leu Glu Glu Lys Xaa Tyr Gln Xaa Ala Glu
20612   115   120   125
W↑→ 20614 Arg Lys Phe Asp Ala Lys Xaa Xaa Arg Asp Arg Ser Xaa Asp Glu Xaa
20615   130   135   140
W↑→ 20617 Lys Lys Thr Xaa Pro Pro Lys Xaa Leu Trp Asn Leu Lys Lys Glu Lys
20618 145   150   155   160
W↑→ 20620 Lys Ser Met Xaa Lys Glu Xaa Glu Lys Glu Thr Glu Glu Arg Arg Lys ← insert
E--> 20621   165   170   175
20623 Ala Leu Glu Met Asp Lys Glu Asn Glu Lys Val Asn Ala Lys Glu Asn
20624   180   185   190
20626 Glu Arg Glu Ile Asn Gln Glu Ser Ala Asn Glu Pro Ser Ser Glu Asn
20627   195   200   205
20629 Thr Pro Leu
E--> 20630   210
20810 (2) INFORMATION FOR SEQ ID NO: 494:
20812   (i) SEQUENCE CHARACTERISTICS:
20813   (A) LENGTH: 78 amino acids
20814   (B) TYPE: amino acid

```
20815            (D) TOPOLOGY: linear
20817       (ii) MOLECULE TYPE: protein
20819      (iii) HYPOTHETICAL: YES
20821       (vi) ORIGINAL SOURCE:
20822            (A) ORGANISM: Helicobacter pylori
20824       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 494:
20826 Met Glu Asn Pro Ser Leu Gly Ser Asn Pro Leu Xaa Gln Lys Ala Met
20827 1               5                  10                  15
20829 Lys Asn Lys Xaa Ile Ser Lys Ser Leu Pro Tyr Try Arg Lys Met Pro    ) invalid
20830             20                  25                  30
20832 Asn Gly Ala Glu Val Tyr Gly Val Xaa Ile Leu Leu Pro Leu Phe Lys
20833             35                  40                  45
20835 Glu Asn Thr Xaa Xaa Trp Trp Gly Val Leu Met Ile Phe Phe Ile Xaa
20836       50                  55                  60
20838 Xaa Xaa Val Met Lys Ser Leu Lys Thr Gly Ala Ile Tyr Phe
20839 65                  70                  75
20996 (2) INFORMATION FOR SEQ ID NO: 497:
20998       (i) SEQUENCE CHARACTERISTICS:            ) 99 shown
20999           (A) LENGTH: 127 amino acids
21000           (B) TYPE: amino acid
21001           (D) TOPOLOGY: linear
21003      (ii) MOLECULE TYPE: protein
21005     (iii) HYPOTHETICAL: YES
21007      (vi) ORIGINAL SOURCE:
21008           (A) ORGANISM: Helicobacter pylori
21010      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 497:
21012 Met Pro Val Ile Ile Gly Tyr Ile Cys Thr Gly Thr Val Leu Ala Ala
21013 1               5                  10                  15
21015 Phe Phe Lys Ile Asn Asp Phe Asn Leu Leu Ser Asp Ile Gly Glu Phe
21016             20                  25                  30
21018 Gly Ile Val Phe Leu Met Phe Lys Lys Gly Ile Glu Phe Asn Phe Asp
21019             35                  40                  45
21021 Lys Leu Lys Ser Ile Lys Gln Glu Val Leu Val Phe Gly Leu Leu Gln
21022       50                  55                  60
21024 Val Val Leu Cys Ala Leu Ile Ala Phe Leu Leu Gly Tyr Phe Val Leu
21025 65                  70                  75                  80
21027 Gly Leu Ser Pro Phe Phe Pro Leu Phe Leu Ala Trp Gly Phe His Ser
21028             85                  90                  95
21030 Leu Gln Pro
22260 (2) INFORMATION FOR SEQ ID NO: 522:
22262       (i) SEQUENCE CHARACTERISTICS:
22263           (A) LENGTH: 107 amino acids
22264           (B) TYPE: amino acid
22265           (D) TOPOLOGY: linear
22267      (ii) MOLECULE TYPE: protein
22269     (iii) HYPOTHETICAL: YES
22271      (vi) ORIGINAL SOURCE:
22272           (A) ORGANISM: Helicobacter pylori
22274      (ix) FEATURE:
```

```
22275              (A) NAME/KEY: misc_feature
22276              (B) LOCATION: 1...107
22277              (D) OTHER INFORMATION: /note= "L-lactate permease"
22279       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 522:
22281  Val Phe Leu Thr Gly Ser Asp Thr Ser Ser Asn Leu Leu Phe Gly Ser
22282  1               5                   10                  15
22284  Leu Gln Met Val Ile Ala Thr Gln Leu Gly Leu Pro Glu Val Leu Phe
22285              20                  25                  30
22287  Leu Ala Ala Asn Thr Ser Gly Gly Val Val Gly Lys Met Ile Ser Pro
22288              35                  40                  45
E-->  22290  Gln Ser Ile Ala Ile Als Cys Ala Ala Val Gly Leu Val Gly Lys Glu   invalid
22291       '  50   '   .    .        55                60
22293  Ser Glu Met Phe Arg Phe Thr Val Lys Tyr Ser Ile Ala Leu Ala Ile
22294  65                  70                  75                  80
22296  Ile Met Gly Ile Val Leu His Ser Tyr Arg Leu Cys Phe Leu Leu Tyr
22297              85                  90                  95
22299  Tyr Ser Ser Tyr Ser Tyr Leu Met Glu Gly Val
22300              100                 105
23164  (2) INFORMATION FOR SEQ ID NO: 540:
23166       (i) SEQUENCE CHARACTERISTICS:                  149 shown
23167              (A) LENGTH: 148 amino acids
23168              (B) TYPE: amino acid
23169              (D) TOPOLOGY: linear
23171       (ii) MOLECULE TYPE: protein
23173      (iii) HYPOTHETICAL: YES
23175       (vi) ORIGINAL SOURCE:
23176              (A) ORGANISM: Helicobacter pylori
23178       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 540:
23180  Met Glu Val Met Asp Glu Ala Leu Leu Arg Ser Lys Arg Phe Asp Arg
23181  1               5                   10                  15
23183  Arg Ile Phe Ile Ser Leu Pro Asp Leu Leu Glu Arg Gln Ser Ile Leu
23184              20                  25                  30
23186  Glu Lys Leu Leu Glu Asn Lys Lys His Ala Leu Asp Tyr Leu Lys Ile
23187              35                  40                  45
23189  Ala Lys Ile Cys Val Gly Phe Ser Gly Ala Met Leu Ala Thr Leu Ile
23190              50                  55                  60
E-->  23192  Aln Glu Ser Ala Leu Asn Ala Leu Lys His Gln Arg Lys Glu Ile Thr   invalid
23193  65                  70                  75                  80
23195  His Gly Asp Ile Leu Glu Val Lys Asp Lys Ile Ala Tyr Gly Lys Lys
23196              85                  90                  95
23198  Lys Pro Gln Thr Leu Asp Glu Asn Gln Lys Glu Leu Val Ala Leu Tyr
23199              100                 105                 110
23201  Gln Ser Ala Lys Ala Leu Ser Ala Tyr Trp Leu Glu Ile Glu Phe Asp
23202              115                 120                 125
23204  Lys Ala Ser Leu Leu Gly Glu Phe Ile Ala Phe Asn Glu Asn Lys Ile
23205       130                 135                 140
23207  His Ala Arg Ala Arg
E-->  23208  145
24988  (2) INFORMATION FOR SEQ ID NO: 578:
```

RAW SEQUENCE LISTING   DATE: 11/27/2002
PATENT APPLICATION: **US/08/487,032B**   TIME: 09:12:45

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

```
 24990       (i) SEQUENCE CHARACTERISTICS:                  →9/ shown
 24991           (A) LENGTH: 90 amino acids
 24992           (B) TYPE: amino acid
 24993           (D) TOPOLOGY: linear
 24995      (ii) MOLECULE TYPE: protein
 24997     (iii) HYPOTHETICAL: YES
 24999      (vi) ORIGINAL SOURCE:
 25000           (A) ORGANISM: Helicobacter pylori
 25002      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 578:
 25004 Met Ser Xaa Xaa Ala Ile Arg Phe Pro Xaa Xaa Leu Phe Ser Tyr Pro
 25005 1               5                   10                  15
 25007 Lys Pro Lys Ile Lys Ala Thr Asn Thr Ser Xaa Thr Val Leu Phe Ala
 25008             20                  25                  30
 25010 Tyr Pro Leu Lys Pro His Glu Met Ala Leu Leu Ala Leu Ala Thr Ser
 25011         35                  40                  45
 25013 Leu Leu Ala Pro Ile Phe Asn Ala Ile His Ser Thr Asn Ala Leu Asn
 25014         50                  55                  60
 25016 Ala Ile Lys Pro Asp Gly Thr Gly Ser Lys Ile Asn Pro Ile Ile Met
 25017 65                  70                  75                  80
 25019 Pro Met Lys Ile Gln Lys Asn Lys Ala Met Arg
 25020                 85                  90   90
 25495 (2) INFORMATION FOR SEQ ID NO: 587:
 25497      (i) SEQUENCE CHARACTERISTICS:
 25498           (A) LENGTH: 182 amino acids
 25499           (B) TYPE: amino acid
 25500           (D) TOPOLOGY: linear
 25502      (ii) MOLECULE TYPE: protein
 25504     (iii) HYPOTHETICAL: YES
 25506      (vi) ORIGINAL SOURCE:
 25507           (A) ORGANISM: Helicobacter pylori
 25509      (ix) FEATURE:
 25510           (A) NAME/KEY: misc_feature
 25511           (B) LOCATION: 1...182
 25512           (D) OTHER INFORMATION: /note= "cation efflux system
 25513 proteins"
 25515      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 587:           → invalid
 25517 Val Ile Val Gly Ala Ile Leu Val Leu Phe Phe Gly Thr Thr Ser Phe
 25518 1.              5                   10                  15
 25520 Ile Asn Thr Pro Val Asp Ala Phe Xaa Asp Ile Ser Pro Thr Gln Val
 25521             20                  25                  30
 25523 Lys Ile Ile Leu Lys Leu Pro Gly Ser Ser Pro Glu Glu Met Glu Asn
 25524             35                  40                  45
 25526 Asn Ile Ala Arg Pro Leu Glu Leu Glu Leu Leu Gly Leu Lys Gly Gln
 25527         50                  55                  60
 25529 Lys Ser Leu Arg Ser Ile Ser Lys Tyr Ser Ile Ser Asp Ile Thr Ile
 25530 65                  70                  75                  80
 25532 Asp Phe Asp Asp Ser Val Asp Ile Tyr Leu Ala Arg Asn Ile Val Asn
 25533             85                  90                  95
 25535 Glu Arg Leu Ser Ser Val Met Lys Asp Leu Pro Val Gly Val Glu Gly
```

```
25536                 100                 105                 110
25538 Gly Met Ala Pro Ile Val Thr Pro Leu Ser Asp Ile Phe Met Phe Thr
25539             115                 120                 125
25541 Ile Asp Gly Asn Ile Thr Glu Ile Glu Lys Arg Gln Leu Leu Asp Phe
25542         130                 135                 140
25544 Val Ile Arg Pro Gln Leu Arg Met Ile Ser Gly Val Ala Asp Val Asn
25545 145                 150                 155                 160
25547 Ser Ile Gly Gly Phe Ser Arg Ala Phe Val Ile Val Pro Asp Phe Asn
25548                 165                 170                 175
25550 Asp Met Ala Arg Leu Gly
25551                 180
31199 (2) INFORMATION FOR SEQ ID NO: 690:
31201      (i) SEQUENCE CHARACTERISTICS:
31202             (A) LENGTH: 144 amino acids
31203             (B) TYPE: amino acid
31204             (D) TOPOLOGY: linear
31206     (ii) MOLECULE TYPE: protein
31208    (iii) HYPOTHETICAL: YES
31210     (vi) ORIGINAL SOURCE:
31211             (A) ORGANISM: Helicobacter pylori
31213     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 690:
31215 Met Lys Asn Leu Arg His Phe Arg Lys Leu Ile Ala Phe Leu Gly Phe
31216 1               5                 10                  15
31218 Ser Pro Leu Leu Leu Gln Ala Asp Met Thr Thr Phe Phe Asn Ser Ile
31219             20                  25                  30
31221 Glu Gln Gln Leu Thr Ser Pro Thr Ala Lys Gly Ile Leu Met Val Ile
31222         35                  40                  45
31224 Phe Leu Gly Leu Ala Ile Phe Ile Trp Lys Asn Leu Asp Arg Trp Lys
31225     50                  55                  60
31227 Glu Ile Leu Met Thr Val Leu Ala Leu Lys Glu Val Pro Met Gln Tyr
31228 65                  70                  75                  80
31230 Phe Ile Pro Ala Ser Asn Leu Lys Glu Ile Ser Ser Lys Glu Lys Phe
31231                 85                  90                  95
31233 Leu Trp Leu Asn Ala Lys Ser Phr Leu Leu Ser Gly Phe Val Pro Phe
31234                 100                 105                 110
31236 Ile Met Ile Pro Trp Leu Asp Ile Leu Asn Ser Phe Val Leu Tyr Val
31237         115                 120                 125
31239 Cys Phe Leu Leu Ile Phe Ser Ile Ala Glu Phe Phe Asp Glu Asp Ile
31240         130                 135                 140
32796 (2) INFORMATION FOR SEQ ID NO: 723:
32798      (i) SEQUENCE CHARACTERISTICS:
32799             (A) LENGTH: 174 amino acids
32800             (B) TYPE: amino acid
32801             (D) TOPOLOGY: linear
32803     (ii) MOLECULE TYPE: protein
32805    (iii) HYPOTHETICAL: YES
32807     (vi) ORIGINAL SOURCE:
32808             (A) ORGANISM: Helicobacter pylori
32810     (ix) FEATURE:
```

E--> (at line 31233)) invalid

RAW SEQUENCE LISTING   DATE: 11/27/2002
PATENT APPLICATION: **US/08/487,032B**   TIME: 09:12:46

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

```
32811              (A) NAME/KEY: misc_feature
32812              (B) LOCATION: 1...174
32813              (D) OTHER INFORMATION: /note= "POTASSIUM/COPPER-TRANSPORTING
32814 ATPASE A"
32816      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 722:    ) invalid
E--> 32818 Met Arg Cys Glu Leu Trp Arg Arg Tyr Gly Gln Htr His Ala Lys Glu
32819 1              5              10              15
32821 Phe gly Pro Tyr Arg Tyr Leu Lys Leu Val Gly Ala Ser Gly Val Gly
32822             20              25              30
32824 Arg Phe Phe Ile Lys Gly Ala Phe Tyr Gly Leu Lys Asn Gly Val Leu
32825        35              40              45
32827 Gly Met Asp Leu Ser Val Ser Phe Gly Ala Leu Ser Ala Phe Val Tyr
32828        50              55              60
32830 Ser Leu Tyr Ala Met Leu Val Ser Gln Glu Thr Tyr Phe Gly Ala Ser
32831 65              70              75              80   ) invalid
E--> 32833 Ser Thr Ile Leu Thr Leu Val Phe Gly Ser Lys Phe Leu Glu Ley Lys
32834             85              90              95
32836 Ala Arg Leu Phe Ala Asn Glu Lys Cys leu Ala Leu Glu Ser His Glu
32837        100             105             110
32839 Ile His Ser Val Ile Val Val Glu Asn Gly Lys Gln Ile Glu Lys His
32840        115             120             125
32842 Pro Lys Asp Val Ala Ile Gly Ser Val Val Trp Val Pro Ser Gly Ala
32843        130             135             140
32845 Lys Ile Ala leu Asp Gly Val Leu Leu Lys Ser Ala Ser Val Asp Ala
32846 145             150             155             160
32848 Ser Leu Ile Ser Gly Glu Phe Lys Pro Leu Glu Ile Gly Gly
32849             165             170
32974 (2) INFORMATION FOR SEQ ID NO: 727:
32976      (i) SEQUENCE CHARACTERISTICS:
32977              (A) LENGTH: 77 amino acids
32978              (B) TYPE: amino acid
32979              (D) TOPOLOGY: linear
32981      (ii) MOLECULE TYPE: protein
32983     (iii) HYPOTHETICAL: YES
32985      (vi) ORIGINAL SOURCE:
32986              (A) ORGANISM: Helicobacter pylori
32988      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 727:
32990 Val His Ser Lys Glu Arg Lys Glu Phe Leu Val Ser Glu Phe Lys Ala
32991 1              5              10              15
32993 Ser Ala Val Glu Met Glu Gly Ala Ser Val Ala Phe Val Cys Gln Lys
E--> 32994 .       20  20                25              30   <--- edit numbering
32996 Phe Gly Val Pro Cys Cys Val Leu Arg Ser Ile Ser Asp Asn Ala Asp
E--> 32997        35              40              45
E--> 32999 Glu Lys Ala Gly Met Ser Phe Asp Glu Phe Leu Glu Lys Ser Ala His
E--> 33000        50              55              60
33002 Thr Ser Ala Lys Phe Leu Lys Ser Met Val Asp Glu Leu
E--> 33003 65              70              75
37478 (2) INFORMATION FOR SEQ ID NO: 821:
37480      (i) SEQUENCE CHARACTERISTICS:
```

```
37481                (A) LENGTH: 84 amino acids
37482                (B) TYPE: amino acid
37483                (D) TOPOLOGY: linear
37485        (ii) MOLECULE TYPE: protein
37487       (iii) HYPOTHETICAL: YES
37489        (vi) ORIGINAL SOURCE:
37490                (A) ORGANISM: Helicobacter pylori
37492        (ix) FEATURE:
37493                (A) NAME/KEY: misc_feature
37494                (B) LOCATION: 1...84
37495                (D) OTHER INFORMATION: /note= "Plasmodium falciparum gametocyte
37496 specific antigen"
```

E--> 37498   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: (824) 821

```
37500 Val Leu Val Val Gly Lys Pro Asn Glu Ser Tyr Ala Asp Thr His Ala
37501 1               5                   10              15
37503 Arg Ile Glu His Phe Ile Lys Leu Val Asp Phe Lys Gly Glu Ile Val
37504             20              25              30
37506 Phe Ile Asn Glu Asp Asn Ser Ser Val Glu Ala Tyr Glu Asn Leu Glu
37507             35              40              45
37509 His Leu Gly Lys Lys Asn Lys Arg Ile Ala Thr Lys Asp Gly Arg Leu
37510       50              55              60
37512 Asp Ser Leu Ser Ala Cys Arg Ile Leu Glu Arg Tyr Cys Gln Gln Val
37513 65              70              75                      80
37515 Leu Lys Lys Gly
38831 (2) INFORMATION FOR SEQ ID NO: 849:
38833        (i) SEQUENCE CHARACTERISTICS:
38834                (A) LENGTH: 46 amino acids
38835                (B) TYPE: amino acid
38836                (D) TOPOLOGY: linear
38838        (ii) MOLECULE TYPE: protein
38840       (iii) HYPOTHETICAL: YES
38842        (vi) ORIGINAL SOURCE:
38843                (A) ORGANISM: Helicobacter pylori
38845        (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 849:
38847 Val Ile Val Gly Val Gly Lys Ser Ala Leu Val Ala Gln Lys Ile Val
38848 1               5                   10              15
38850 Ala Ser Met Leu Ser Thr Gly Asn Arg Ser Ala Phe Leu His Pro Thr
38851             20              25              30
```

E--> 38853 Glu Ala Met His Gly Asp Leu Gly Met Gly Lys Lys Thr (Mer) invalid
```
38854             35              40              45
39047 (2) INFORMATION FOR SEQ ID NO: 854:
39049        (i) SEQUENCE CHARACTERISTICS:
39050                (A) LENGTH: 153 amino acids
39051                (B) TYPE: amino acid
39052                (D) TOPOLOGY: linear
39054        (ii) MOLECULE TYPE: protein
39056       (iii) HYPOTHETICAL: YES
39058        (vi) ORIGINAL SOURCE:
39059                (A) ORGANISM: Helicobacter pylori
```

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

```
39061     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 854:
39063 Met Asp Ala Leu Glu Ile Thr Gln Lys Leu Ile Ser Tyr Pro Thr Ile
39064 1               5                   10                  15
39066 Thr Pro Lys Glu Cys Gly Ile Phe Glu Tyr Ile Lys Ser Leu Phe Pro
39067             20                  25                  30
39069 Ala Phe Lys Thr Leu Glu Cys Glu Lys Asn Gly Val Lys Asn Leu Phe
39070         35                  40                  45
39072 Leu Tyr Arg Ile Phe Asn Pro Leu Lys Lys His Ala Glu Lys Glu His
39073     50                  55                  60
39075 Ala Lys Glu Lys His Val Lys Glu Asn Val Xaa Pro Leu His Phe Cys
39076 65              70                  75                  80
39078 Leu Gln Gly Ile Leu Unk Ser Cys Leu Leu Gly Xaa Xaa Ala Xaa Asp
39079             85                  90                  95
39081 Ser Phe Xaa Xaa Ile Ile Lys Glu Gly Phe Leu Tyr Gly Arg Gly Ala
39082         100             105                 110
39084 Gln Asp Met Lys Gly Gly Val Gly Xaa Phe Leu Gly Ala Xaa Xaa Asn
39085         115             120             125
39087 Phe Asn Xaa Lys Xaa Xaa Phe Xaa Phe Leu Phe Tyr Leu Thr Ser Asp
39088     130             135             140
39090 Glu Glu Gly Thr Arg Xaa Phe Xaa His
39091 145             150
43542 (2) INFORMATION FOR SEQ ID NO: 941:
43544     (i) SEQUENCE CHARACTERISTICS:
43545         (A) LENGTH: 55 amino acids
43546         (B) TYPE: amino acid
43547         (D) TOPOLOGY: linear
43549    (ii) MOLECULE TYPE: protein
43551   (iii) HYPOTHETICAL: YES
43553    (vi) ORIGINAL SOURCE:
43554         (A) ORGANISM: Helicobacter pylori
43556    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 941:
43558 Met Trp Leu Ser Glu His Phe Ala Ala Lys Gly Gly Asn Pro Leu Phe
43559 1               5                   10                  15
43561 Ala Pro Tyr Tyr Leu Gln Asp Asn Pro Thr Glu His Ile Val Thr Leu
43562             20                  25                  30
43564 Met Lys Asp Ile Thr Ser Ala Leu Gly Met Leu Ser Asn Ser Asn Leu
43565         35                  40                  45
43567 Lys Asn Asn Ser Thr Asp Val Leu Gln Leu Asn Thr Tyr Thr Gln Gln
43568     50                  55                  60
43570 Met Ser Arg Leu Ala Lys Leu Ser Asn Phe Ala Ser Phe Asp Ser Thr
43571 65              70                  75                  80
43573 Asp Phe Ser Glu Arg Leu Ser Ser Leu Lys Asn Gln Arg Phe Ala Asp
43574             85                  90                  95
43576 Ala Val Pro Asn Ala Met Asp Val Ile Leu Lys Tyr Ser Gln Arg Asp
43577             100             105                 110
43579 Lys Leu Lys Asn Asn Leu Trp Ala Thr Gly Val Gly Gly Val Ser Phe
43580         115             120             125
43582 Val Glu Asn Gly Thr Gly Thr Leu Tyr Gly Val Asn Val Gly Tyr Asp
43583     130             135             140
```

(handwritten annotations in margin)

E--> (line 39078)

W--> (line 39081)

W--> (line 39084)

W--> (line 39087)

W--> (line 39090)

C--> (line 43542)

)Invalid if this denotes "unknown", use Xaa

)194 (see p.11)

RAW SEQUENCE LISTING   DATE: 11/27/2002
PATENT APPLICATION: **US/08/487,032B**   TIME: 09:12:47

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

```
43585 Arg Phe Val Arg Gly Val Ile Val Gly Gly Tyr Ala Ala Tyr Gly Tyr
43586 145              150              155          160
43588 Ser Gly Phe Tyr Glu Arg Ile Thr Ser Ser Lys Ser Asp Asn Val Asp
43589              165              170              175
43591 Val Gly Met Tyr Ala Arg Ala Phe Ile Lys Lys Ser Glu Leu Thr Phe
43592          180              185              190
```
E--> **43594 Arg Arg**

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 11/27/2002
PATENT APPLICATION: **US/08/487,032B**   TIME: 09:12:48

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

<u>Invalid Line Length:</u>
The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:666; Line(s) 30012

VERIFICATION SUMMARY   DATE: 11/27/2002
PATENT APPLICATION: **US/08/487,032B**   TIME: 09:12:48

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

```
L:28 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:29 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:280 M:111 C: (47) String data converted to upper case,
M:111 Repeated in SeqNo=6
L:1844 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:2860 M:111 C: (47) String data converted to upper case,
M:111 Repeated in SeqNo=74
L:6118 M:220 C: Keyword misspelled or invalid format, [(xi.) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:6191 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:6390 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:6666 M:111 C: (47) String data converted to upper case,
L:7487 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:8652 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:8802 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:9469 M:111 C: (47) String data converted to upper case,
L:10267 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:13814 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:13989 M:111 C: (47) String data converted to upper case,
L:15383 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:387 after pos.:0
L:15386 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:387 after pos.:16
L:15488 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:389 after pos.:0
L:15588 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:391 after pos.:48
L:15614 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:392 after pos.:16
L:15647 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:392 after pos.:192
L:15862 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:395 after pos.:64
L:16101 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:399 after pos.:240
L:16110 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:399 after pos.:288
L:16286 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:402 after pos.:256
L:16330 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:403 after pos.:128
L:16336 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:403 after pos.:160
L:16532 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:407 after pos.:0
L:16767 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:412 after pos.:48
L:16770 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:412 after pos.:64
L:17183 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:421 after pos.:80
L:17212 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:422 after pos.:32
L:17268 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:423 after pos.:128
L:17431 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:427 after pos.:192
L:17608 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:432 after pos.:48
L:17640 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:433 after pos.:32
L:17646 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:433 after pos.:64
L:17788 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:437 after pos.:0
L:18347 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:18392 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:449 after pos.:48
L:18659 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:454 after pos.:112
L:18668 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:454 after pos.:160
L:18807 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:458 after pos.:48
L:18817 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:458
L:18994 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:461 after pos.:160
```

VERIFICATION SUMMARY   DATE: 11/27/2002
PATENT APPLICATION: US/08/487,032B   TIME: 09:12:48

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

```
L:19254 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:467 after pos.:0
L:19550 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:473
L:20611 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:490 after pos.:112
L:20614 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:490 after pos.:128
L:20617 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:490 after pos.:144
L:20620 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:490 after pos.:160
L:20621 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:490
L:20630 M:203 E: No. of Seq. differs, LENGTH:Input:166 Found:211 SEQ:490
L:20826 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:494 after pos.:0
L:20829 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:494 after pos.:16
L:20829 M:330 E: (2) Invalid Amino Acid Designator, 1
L:20832 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:494 after pos.:32
L:20835 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:494 after pos.:48
L:20838 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:494 after pos.:64
L:21030 M:203 E: No. of Seq. differs, LENGTH:Input:127 Found:99 SEQ:497
L:21122 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:499 after pos.:0
L:21128 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:499 after pos.:32
L:21131 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:499 after pos.:48
L:21134 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:499 after pos.:64
L:21276 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:502 after pos.:160
L:21279 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:502 after pos.:176
L:21553 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:508 after pos.:16
L:21593 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:509 after pos.:32
L:21608 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:509 after pos.:112
L:22167 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:519 after pos.:64
L:22173 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:519 after pos.:96
L:22290 M:330 E: (2) Invalid Amino Acid Designator, 1
L:22320 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:22333 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:523 after pos.:48
L:22436 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:23192 M:330 E: (2) Invalid Amino Acid Designator, 1
L:23208 M:203 E: No. of Seq. differs, LENGTH:Input:148 Found:149 SEQ:540
L:23569 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:24040 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:25020 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:578
L:25020 M:203 E: No. of Seq. differs, LENGTH:Input:90 Found:91 SEQ:578
L:25517 M:333 E: Wrong sequence grouping, Amino acids not in groups!
L:25517 M:330 E: (2) Invalid Amino Acid Designator, 1
L:26015 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:26337 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:26504 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:29075 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:31178 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:31233 M:330 E: (2) Invalid Amino Acid Designator, 1
L:31928 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:32818 M:330 E: (2) Invalid Amino Acid Designator, 1
L:32833 M:330 E: (2) Invalid Amino Acid Designator, 1
L:32994 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:727
M:332 Repeated in SeqNo=727
```

VERIFICATION SUMMARY   DATE: 11/27/2002
PATENT APPLICATION: **US/08/487,032B**   TIME: 09:12:48

Input Set : **D:\Seqlistcorr2.txt**
Output Set: **N:\CRF4\11272002\H487032B.raw**

```
L:37498 M:212 E: (34) Invalid or duplicate Sequence ID Number, Value=[824:]
L:38853 M:330 E: (2) Invalid Amino Acid Designator, 1
L:38926 M:220 C: Keyword misspelled or invalid format, [(xi) SEQUENCE DESCRIPTION: SEQ ID NO:]
L:39078 M:330 E: (2) Invalid Amino Acid Designator, 1
L:40255 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:41817 M:220 C: Keyword misspelled or invalid format, [(D) TOPOLOGY:]
L:41817 M:246 W: Invalid value of Alpha Sequence Header Field, [TOPOLOGY:], SeqNo=906
L:43424 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:43513 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:43542 M:220 C: Keyword misspelled or invalid format, [(2) INFORMATION FOR SEQ ID NO:]
L:43594 M:203 E: No. of Seq. differs, LENGTH:Input:55 Found:194 SEQ:941
```